AFCRL-63-331

INTEGRATED LOGIC NETS

Prepared by

S. R. Hofstein and F. P. Heiman

RADIO CORPORATION OF AMERICA
RCA LABORATORIES
PRINCETON, NEW JERSEY

SCIENTIFIC REPORT NO. 2

Contract No. AF19(604)-8836
Project No. 4641
Task No. 464104

JULY 16, 1963

Prepared for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

ABSTRACT

The device discussed in this report represents a significant departure from the conventional unipolar transistor in that the reverse-biased p-n junction has been superseded by a metal-oxide-semiconductor control structure. This structure can be used to enhance as well as deplete the charge near the surface of the semiconductor. A simple model is proposed and the basic transistor current-voltage relationships are derived for the case of a thick oxide and shallow conducting channel. This case is in contrast to the p-n junction-type described by Shockley, the latter being analogous to a very thin oxide and a deep, uniformly doped channel.

A more detailed model is subsequently proposed in order to explain some experimentally observed anomalies for units not adequately described by the simpler model. In particular, the usual approximation of constant current in the saturation region is abandoned and consideration is given to the behavior of the drain resistance in this region. It is found that the relations predicting and describing this behavior are closely analogous to similar relations for vacuum tubes. Experimental data from units to which this model may be applied have shown close agreement with the theoretical predictions.

Breakdown in the channel is also investigated and found to be a double-valued function of the gate voltage. For low gate voltages the breakdown is induced directly by the drain-to-source field and occurs across the constricted portion of the channel near the drain. For higher gate voltages carrier generation is induced directly by the gate-to-drain field with the source-to-drain sweep field acting to remove the impact-ionized and/or field-emitted carriers.

Brief consideration is also given to the question of the validity of using Boltzmann or Fermi-Dirac statistics in devices which are not necessarily in thermal equilibrium. The use of these statistics to predict the formation of a gate-field induced inversion layer at the oxide-silicon interface in typical depletion-type transistors is found to be unjustified. This is due to generation and flow-rate limitations on the source of the carriers which form this layer, and contrasts with the case of the induced channel-type unit for which the use of these statistics appears justified. Experimental evidence regarding saturation and complete pinch-off of drain current has indicated that the inversion layer does, in fact, fail to exist, as predicted, for the depletion-type transistor.

Units fabricated to date have the following typical characteristics: Input Impedance – 7 $\mu\mu f$, 10^{+15} ohms; Transconductance – 2000 μmhos; Cut-Off Bias – –7 volts (Depletion Unit); and Rise Time – 10 nanoseconds. The yield of units has averaged over 95% on recently fabricated wafers, indicating great promise for integrated electronics applications.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

INTRODUCTION

As described by Shockley,[1] the unipolar field-effect transistor utilized the depletion region of a reverse-biased p-n junction to control the effective cross section, and hence the conductance, of a bar of semiconductor material. This device is illustrated in Fig. 1. The ohmic contacts are conventionally referred to as the source and drain rather than emitter and collector so as to emphasize the fact that they, in essence, inject and remove only majority carriers. The conductive region between the ohmic source and drain contact is, again by convention, referred to as the channel, with the reverse-biased p-n junction space-charge control electrode acting as a "gate." The

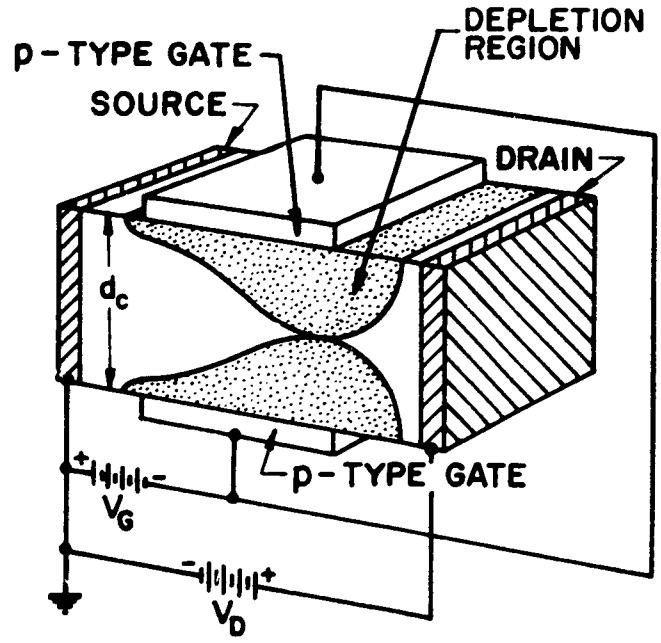


Fig. 1. Basic unipolar transistor structure.

theoretical characteristics of this device derived by Shockley[1] and others[2] have since been experimentally confirmed.[2] The device discussed in this report represents a significant departure from the Shockley unit in that the reverse-biased p-n junction has been superseded by a metal-insulator-semiconductor control structure.[3] In sharp contrast to the p-n junction, this structure can be used to enhance as well as deplete charge near the surface of the semiconductor; this results in a field-effect device possessing a significantly increased versatility.

The concept of using an external electric field normal to the surface of a semiconductor to control the carrier density near the surface is discussed by Shockley and Pearson.[4] They

placed a thin insulating strip between the semiconductor and an evaporated metal film forming a parallel-plate capacitor. The change in conductance of the semiconductor was measured as a function of the voltage applied across the capacitor. The results of this experiment showed that roughly 10% of the excess charge placed in the germanium was mobile. The rest was postulated to reside in bound states on the surface of the semiconductor. It has been experimentally observed that surface states (both acceptor and donor types) are distributed in energy throughout the forbidden gap and are further characterized by a delay in response, or time constant.[5] These states have also been found to be highly dependent upon surface conditions, including the effects of adsorbed gases, surface strains, mechanical damage, etc.

Many characteristics of the ideal Metal-Oxide-Semiconductor (MOS) structure have been derived elsewhere.[5,6] Experimental measurements have confirmed the general predictions made for these structures[5] and have, in addition, provided some specific information on the density, energy levels, and time constants of the surface states at a silicon-silicon dioxide interface.

It has been reported[6,7] that silicon dioxide which is thermally grown on the surface of a silicon wafer has the property of passivating the surface, or greatly decreasing the density of deep surface traps. This decrease has been found sufficient to reduce the effects of these states on device characteristics to the point where they may be neglected.

This report* is divided into five sections: Section I derives the first order theoretical device parameters; Section II is a discussion of the experimentally observed parameters; Section III concerns itself with several aspects of the physical theory of MOS-structure devices necessary to understand some significant deviations from the first order theory of Section I; Section IV contains a discussion of experimental results pertaining to the predictions of Section III, and Section V deals with the application of the device to integrated circuitry.

A minimum of mathematics has been used in Section III in order to emphasize the physics rather than the algebra of the phenomena involved.

*The work described in this report was first presented at the Electron Devices Meeting, Washington, D. C., October 25-27, 1962, in a paper entitled "The Insulated-Gate Field-Effect Transistor," by F. P. Heiman and S. R. Hofstein.

I. THEORY OF TRANSISTOR OPERATION

A. BASIC STRUCTURE

The basic structure of the insulated gate device we have fabricated is illustrated in Fig. 2. This is a planar device, with the substrate passive and acting as a support.* The channel, source and drain contacts, and control electrode are fabricated using conventional diffusion, photoresist, and vacuum-evaporation techniques. It might be noted that this device differs from the Thin-Film Transistor (TFT) described by Weimer[8] in that the conduction in Weimer's device takes place in an evaporated thin film rather than a layer genetically derived from the substrate.

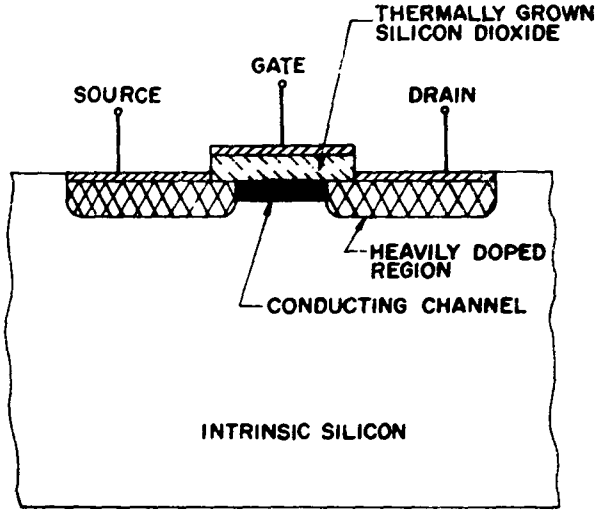


Fig. 2. Insulated-gate field-effect transistor structure.

B. DEPLETION-TYPE TRANSISTOR

Referring to Fig. 2, consider first a transistor fabricated with the channel of the same conductivity type as the source and drain contacts; this may be operated in either the *depletion or enhancement* modes.

Depletion Mode: To operate in this mode, the gate is reverse-biased so that carriers are depleted from the channel. Therefore, maximum channel current flows in this mode of operation for zero gate bias. This is analogous to the operation of the Shockley unit.

*In practice, the substrate may be modified to form a junction with the channel and a bias voltage may then be applied to the substrate. However, for the following analysis, a passive substrate will be assumed.

Enhancement Mode: Here, the gate is forward-biased so that carriers are drawn into the channel; minimum channel current flows in this mode of operation for zero gate bias. In contrast to the Shockley unit, no gate current flows due to the insulating silicon dioxide layer.

C. INDUCED CHANNEL-TYPE TRANSISTOR

This unit is fabricated with source and drain contacts of opposite conductivity type from the channel. Back-to-back diodes are formed between source and drain contacts and the channel current is essentially zero for zero gate bias. To simplify the following discussion, a p-type channel with n-type source and drain contacts will be assumed. If a positive voltage is applied to the gate, holes will be depleted from the surface and a further increase in bias will produce an accumulation of electrons at the surface; the surface channel goes from p-type, through intrinsic, to an inverted n-type layer, at which point ohmic conduction from source to drain commences. Enhancement-mode operation is obtained for a further increase in gate bias.

In the following analysis a characteristic parameter for the unit will be called the pinch-off voltage V_p. This is defined as the potential difference across the oxide (at any point) which will just cause the mobile charge concentration in the channel (at that point) to go to zero. It may be positive (electron conduction induced-channel device), negative (electron conduction depletion device), or zero. Also, the oxide is assumed to be much thicker than the channel depth so that almost all of the control voltage appears across the oxide and very little across the charged semiconductor surface. With this approximation, the channel is essentially a surface sheet, and the electric field in the oxide is normal to the semiconductor surface.

With reference to Fig. 3, let the drain current be I_d, with the external source connection grounded and with the external drain and gate electrodes biased at V_D and V_G, respectively. The internal source and drain connections differ in potential from the accessible terminals by the potential drop across the parasitic series resistances R_S and R_D, respectively.

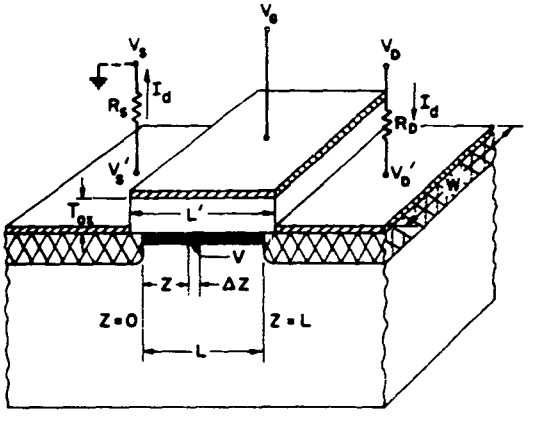


Fig. 3. Schematic representation used in calculations.

D. ANALYSIS OF AN ELECTRON-CONDUCTION DEVICE

An electron conduction device is considered and $n+$ source and drain contacts are employed. The potential at some point, Z, in the channel is given by $V(Z)$, where $0 \leq V(Z) \leq V_D$. With a thick oxide implicitly assumed, one has

$$E_{ox}(Z) = \frac{V_G - V(Z)}{T_{ox}} \tag{1}$$

and

$$\sigma_c(Z) = \epsilon_{ox} E_{ox}(Z) = \frac{\epsilon_{ox}}{T_{ox}} \left[V_G - V(Z)\right], \tag{2}$$

where σ_c is the induced surface-charge density in the channel at point Z, ϵ_{ox} is the dielectric constant of the oxide, and T_{ox} is the oxide thickness. Not all of this charge may be mobile. By the definition of V_P, there will be mobile charge at the point Z only if $V_G - V(Z) \geq V_P$. Thus, we have for the mobile charge density, σ_m, at any point,*

$$\sigma_m(Z) = \frac{\epsilon_{ox}}{T_{ox}} \{[V_G - V(Z)] - V_P\}: \qquad \text{for } V_G - V(Z) > V_P$$

$$\sigma_m(Z) = 0, \qquad \text{for } V_G - V(Z) \leq V_P. \tag{3}$$

The conductance, $G(Z)$ of the infinitesimal channel section of width W and of length ΔZ is given by

$$G(Z) = \frac{\sigma_m(Z)\mu W}{\Delta Z} \tag{4}$$

where μ is the carrier mobility, assumed constant in this analysis. Ohm's law yields the channel current:

$$I_d = G(Z)\Delta V = \sigma_m(Z)\mu W \frac{\Delta V}{\Delta F}\ , \tag{5}$$

whence

$$I_d = \frac{\epsilon_{ox}\mu W}{T_{ox}} [V_G - V_P - V(Z)] \frac{\Delta V}{\Delta Z}\ . \tag{6}$$

An integration from one end of the channel to the other yields:

$$I_d \int_o^L dZ = \frac{\epsilon_{ox}\mu W}{T_{ox}} \int_{V_S'}^{V_D'} [V_G - V(Z) - V_P] dV\ , \tag{7}$$

*The presence of intrinsic channel charge to be depleted before reaching the onset of channel conduction merely offsets the relations involving gate voltage by an amount V_P.

$$I_d = \frac{\epsilon_{ox}\mu W}{LT_{ox}} \left[(V_G - V_P)(V_D' - V_S') - \tfrac{1}{2}(V_D'^2 - V_S'^2)\right] \tag{8}$$

The two auxiliary equations needed to eliminate V_D' and V_S' are

$$V_D' = V_D - I_d R_D , \qquad V_S' = I_d R_S , \tag{9}$$

and substitution into Eq. (8) yields the desired set of drain characteristics for this device. However, this solution is valid only for $V_G - V_D' \geq V_P$, for, when the difference between gate and internal drain voltage is less than V_P, the channel is pinched-off near the drain and the limits of integration used in Eq. (7) are no longer valid. As in the Shockley unit, the current remains essentially constant[1] for larger values of drain voltage and will be designated by I_{ds}. Imposing the condition that $V_D' = V_G - V_P$ yields I_{ds} from Eqs. (8) and (9).

$$I_{ds} = \beta \frac{(V_G - V_P)^2}{1 + \beta R_S (V_G - V_P) + \sqrt{1 + 2\beta R_S (V_G - V_P)}} \tag{10}$$

and

$$\beta \equiv \frac{\epsilon_{ox}\mu W}{LT_{ox}} \tag{11}$$

It should be noted that for small values of series source resistance, the transfer function is essentially parabolic yielding a square-law device. The drain current beyond saturation is essentially independent of the parasitic series drain resistance R_D. This resistance merely increases the drain voltage at which the drain current saturates.

Differentiating Eq. (10) yields the transconductance of the device as shown below,

$$g_m = \frac{\partial I_{ds}}{\partial V_G} = (2\beta)\left[\frac{(V_G - V_P)}{1 + 2\beta R_S (V_G - V_P) + \sqrt{1 + 2\beta R_S (V_G - V_P)}}\right] \tag{12}$$

The total input capacitance, C_{in}, is simply the parallel plate capacitor formed by the gate electrode of area, $L'W$, separated from the semiconductor by spacing, T_{ox}. In practice, $L' \approx L$ and the g_m/C_{in} ratio is a good measure of the frequency response of this device.

$$g_m/C_{in} = \left(\frac{2\mu}{L^2}\right)\left(\frac{(V_G - V_P)}{1 + 2\beta R_S (V_G - V_P) + \sqrt{1 + 2\beta R_S (V_G - V_P)}}\right). \tag{13}$$

II. EXPERIMENTAL PARAMETERS

We have experimentally observed the appearance of a thin n-type skin in the surface of p-type silicon after thermal oxidation. Using this shallow inversion layer as a channel, a field-effect transistor capable of operation in the depletion mode was fabricated on 1000 ohm-cm p-type silicon, using the thermally produced n-type inversion layer for the channel. The channel length, L, was 0.5 mil, the channel width, W, was 50 mils, and the oxide thickness, T_{ox}, was 2000 Å. The family of drain characteristics is shown in Fig. 4. A small-signal transductance of 2000 μmhos is indicated for the device which is operated in both the depletion and enhancement modes. If Eq. (10) is normalized to the saturation drain current for $V_G = 0$ and expressed as a function of normalized gate voltage, $V' = V_G/V_P$, the result is

$$\frac{I_{ds}}{I_{ds}|_{V_G=0}} \equiv I'_{ds} = (1 - V')^2 \left(\frac{1 + (K/2) + \sqrt{1+K}}{1 + (K/2)(1 - V') - \sqrt{1 + K(1 - V')}} \right), \tag{14}$$

where,

$$K \equiv -2\beta R_S V_P \ . \tag{15}$$

Equation (14) is plotted in Fig. 5 (for operation in the depletion mode) for several values of K; the experimental points on this plot were taken from the drain characteristics shown in Fig. 4 at a drain voltage of 16 volts. The close agreement suggests that $K = 2$ for this device, and

$$-2\beta R_S V_P = 2,$$

$$R_S = -\frac{1}{\beta V_P} \ . \tag{16}$$

A simple algebraic manipulation shows that $-\beta V_P$ is actually the channel conductance at zero gate bias for a depletion device; thus, the parasitic source resistance, R_S, is equal to the zero-bias channel resistance for this device.

By use of a mercury relay pulser and sampling oscilloscope, the pulse response of this unit was measured. The circuit used and waveforms obtained are shown in Fig. 6; a rise time (10% - 90%) of 10 nanoseconds is indicated. No storage or transit delays appear to be present, and an improvement in dimensions should yield a corresponding improvement in response time.

The induced channel unit is desirable for use in logical switching circuits since the output voltage of the transistor is of the correct polarity to make direct-coupled transistor logic feasible, without shifting voltage levels between stages. The device, the drain characteristics of which are shown in Fig. 7, was fabricated on 5 ohm-cm p-type silicon; the inversion layer

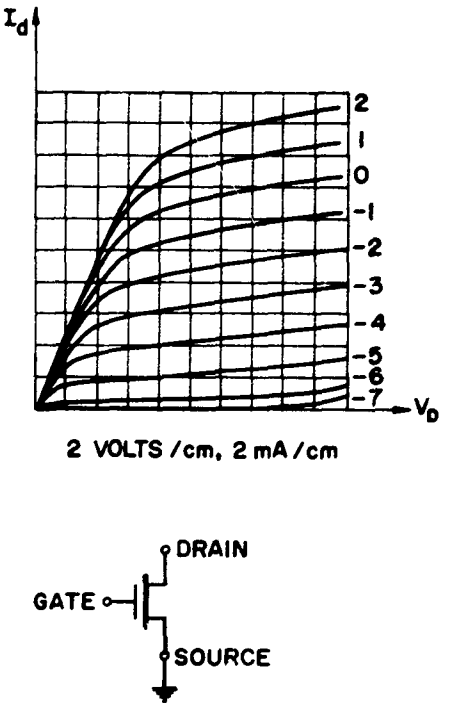


Fig. 4. Drain characteristics of experimental transistor.

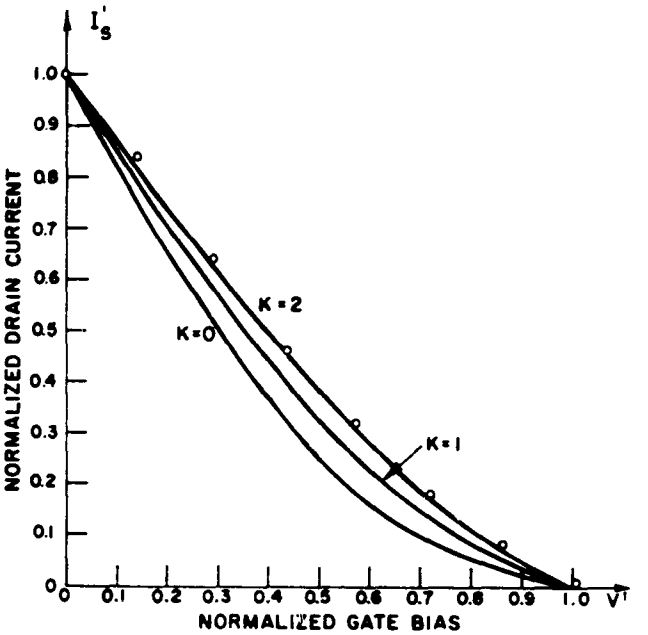
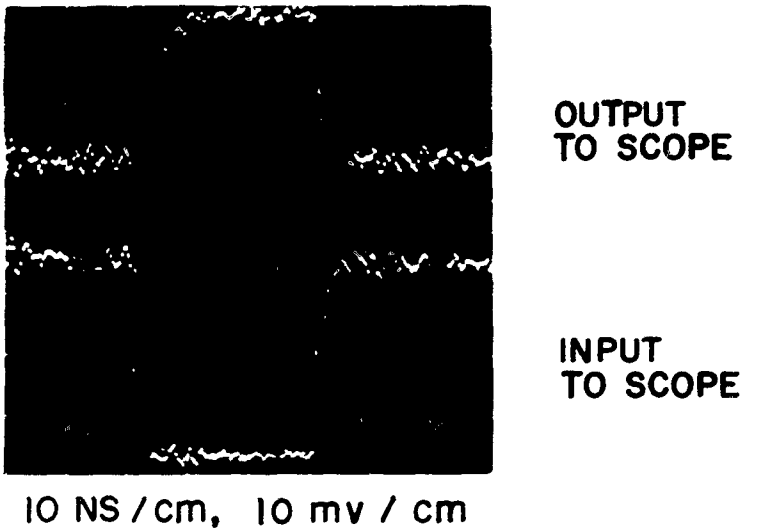


Fig. 5. Variation in drain current with gate bias.



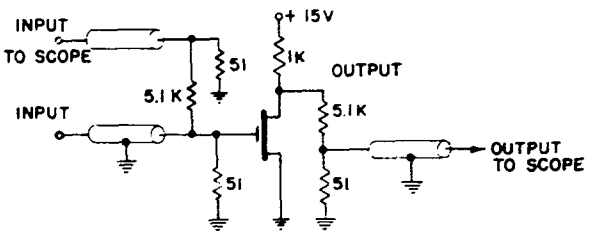


Fig. 6. Pulse response of transistor.

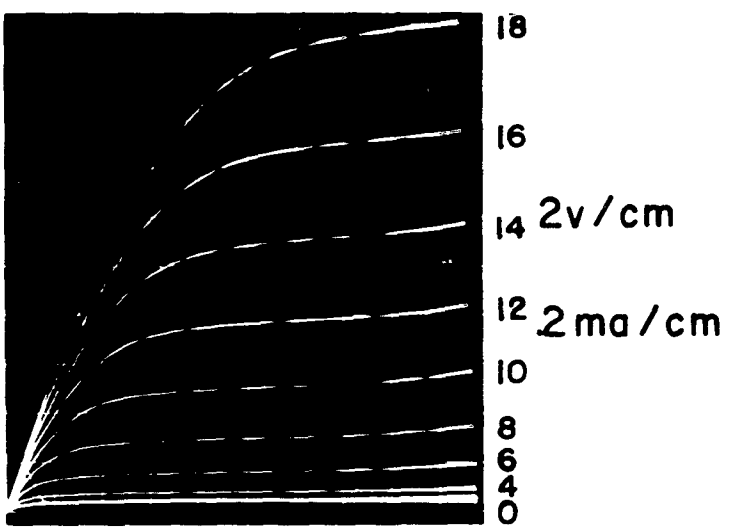


Fig. 7. Induced channel n-p-n field-effect transistor.

usually produced during thermal oxidation was compensated by a pre-oxidation p-type diffusion. The channel length, L, of this device is 0.5 mil, its width, W, is 5 mils, and the oxide thickness, T_{ox}, is 1000 Å. The effective pinch-off voltage for this device is approximately +2 volts. Figure 8 compares the transfer characteristics of the depletion unit with the induced channel unit. The drain current scale is normalized to unit channel width.

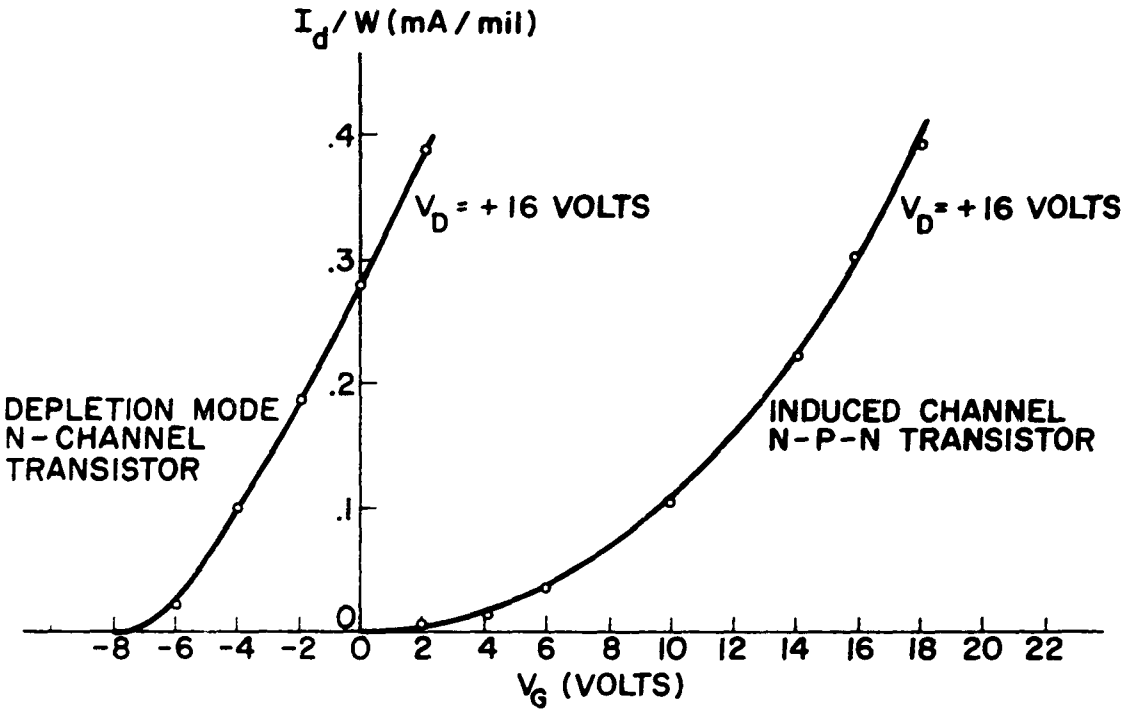


Fig. 8. Comparison of transfer characteristics for field-effect transistors.

Figure 9 shows the layout of induced channel field-effect transistors arranged in a ladder pattern. The interconnection of many devices in an integrated circuit is greatly facilitated with this geometry. The packing density shown in Fig. 9 is 2200 transistors per square inch; this allows ample passive surface area for heat dissipation, interconnection wiring and dicing.

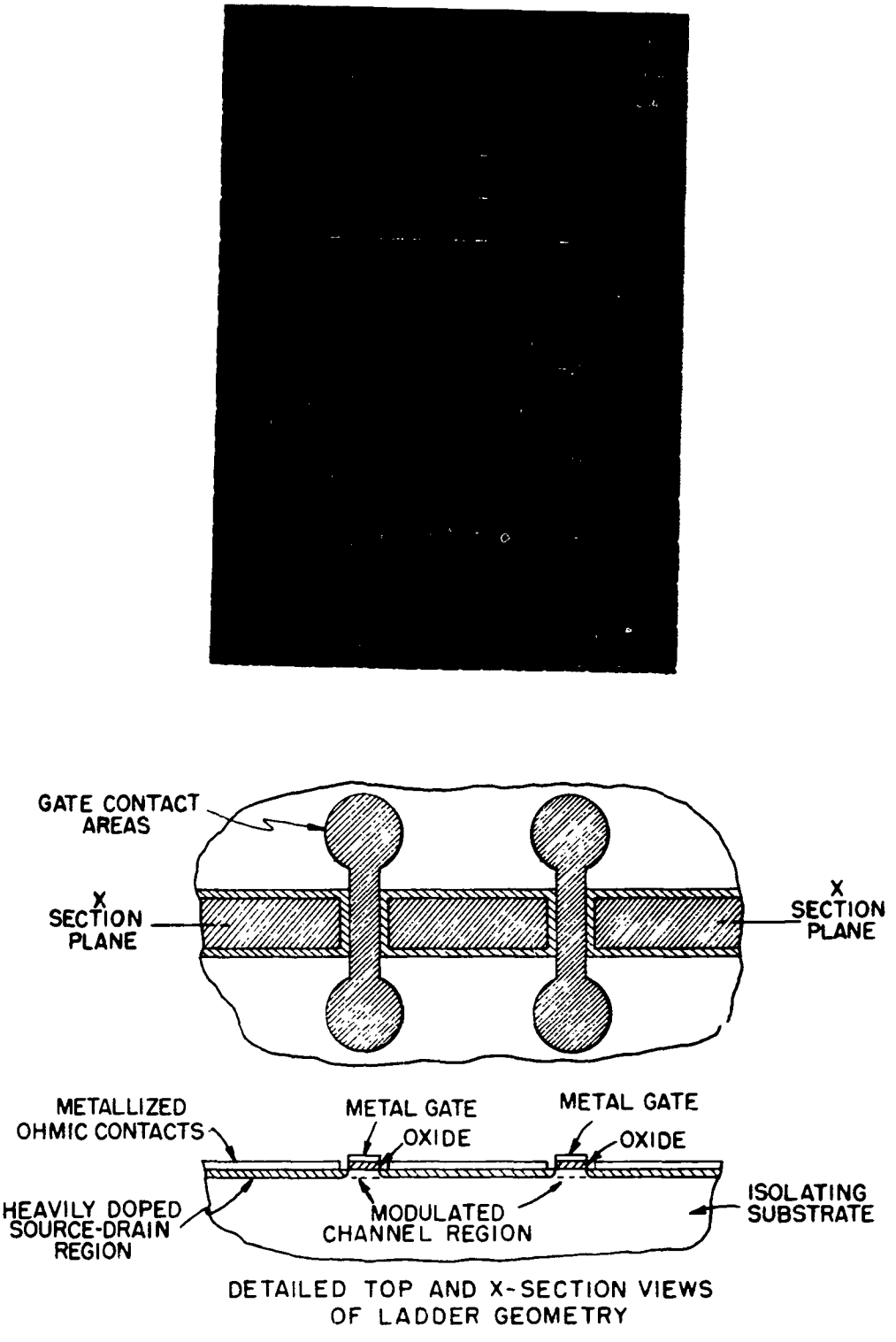


Fig. 9. Ladder geometry layout of field-effect transistor.

III. SOME ADDITIONAL ASPECTS OF THE PHYSICS OF MOS DEVICES

Experimentally observed characteristics of the units described in this report have possessed anomalies not explicable in terms of the simpler models.[1,2,5] These anomalies may be understood if the model is extended to include: A. The effect of the failure of the Shockley condition (channel length to depth ratio less than two), and B. The general problem of thermal equilibrium in three-terminal and non-ideal two-terminal MOS field-effect devices.

A. THEORY OF THE DRAIN SATURATION RESISTANCE

For convenience, an n-type channel and n-type source-drain contacts will be assumed in the following discussion. However, the discussion may be readily extended to include units which operate in the induced channel mode of operation without any change in the physical principles involved.

The pentode-like characteristics of the field-effect transistor, in regard to the saturation of the drain current, is basically a geometry-dependent effect. A good qualitative understanding may be obtained from the following: As the voltage at the drain end of the channel exceeds $V_G - V_P$, the depletion region at that point extends completely across the channel. The channel may now be thought of as consisting of three regions: Region 1 – the section from the source up to the point where the channel voltage equals V_G*; Region 2 – from the V_G point to the point where the channel voltage is $V_G - V_P$; and finally, Region 3 – from the $V_G - V_P$ point to the drain contact. It should be noted that in Region 1 there is an enhancement of the channel charge, in Region 2 there is a partial depletion of channel charge, and in Region 3 the depletion region extends completely across the channel. Hence, in Regions 1 and 2 the current flow is essentially ohmic whereas in Region 3 it is space-charge limited.

Referring now to the Shockley-type (depletion mode) model of Fig. 1, the channel may be divided into a Region 2 and a Region 3** in accordance with the preceding discussion. The penetration of the drain field into Regions 2 and 3 is an exponentially decaying function with a "length constant" equal to the total channel depth, d_c.[1] This result hinges on the assumption that the gates are so highly doped as to be essentially metallic. If channel depth is very much smaller than channel length, then this penetration will be very small and the channel shape will be only slightly affected by variations in drain voltage. Hence, the resistance of Region 2 is nearly constant as V_D varies past the saturation value $V_G - V_P$. Since the voltage across Region 2 is fixed (by definition) at $V_G - V_P$, the current through this region (and by continuity through

*Region 1 exists only for $V_G > 0$.

**Enhancement of channel charge, and hence a Region 1 can exist only for an insulated gate so that $V_G > 0$ does not result in a large gate current.

Region 3) is essentially independent of changes in V_D for V_D greater than $V_G - V_P$. Therefore, for this model, one would expect an extremely high saturation drain resistance. In the fabricated structure (Fig. 2) there are several additional mechanisms which act to reduce this resistance to a value below that predicted by the simple model.

1. Effective Failure of the Shockley Condition

The channel length-to-depth ratio for the units under discussion is typically of the order of 100. However, penetration of the drain field into the channel region is a function of the effective spacing between the metal gate and the "substrate gate." (Since the substrate is slightly p-type, it acts as a control electrode as in the Shockley unit.) This spacing includes the actual doped channel depth plus the space-charge region between the channel and substrate. For a sufficiently low doped substrate, this spacing may be as large as (or even exceed) the channel length as the drain voltage rises into the saturation region.

Hence, the Shockley condition no longer holds and a new model must be postulated. This model is shown in Fig. 10. The basic assumption is that the substrate is sufficiently low-doped so that, for dimensions of the order of the channel length, the contribution of the substrate space charge is negligible. In other words, the substrate is treated as an insulator and Laplace's equation rather than Poisson's equation is used to determine the substrate fields. The drain electrode now acts both as a collector of channel carriers and as a somewhat inefficient gate.

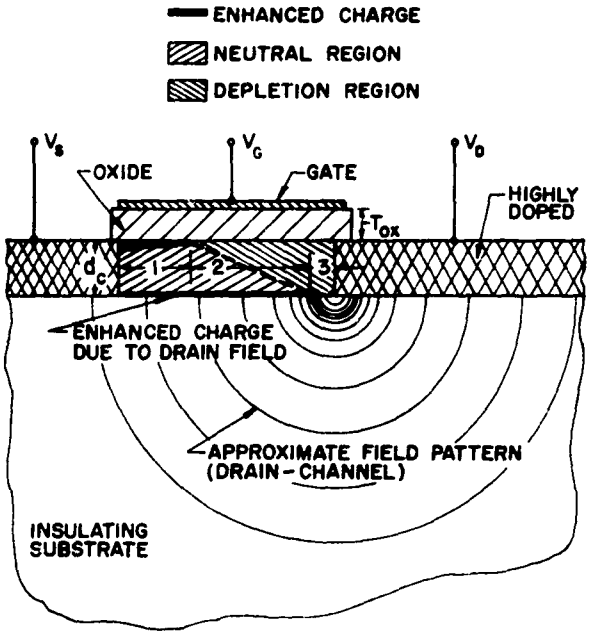


Fig. 10. Model for calculating the saturation drain resistance.

If the special case of oxide thickness >> channel depth is considered, the saturation drain resistance may be directly related to the input transconductance of the device.*

Defining $$\frac{1}{r_{ds}} = g_{ds} \equiv \left.\frac{\partial I_{ds}}{\partial V_D}\right|_{V_G} , \qquad \mu_A \equiv \left.\frac{\partial V_D}{\partial V_G}\right|_{I_d} ,$$

and

$$g_m \equiv \left.\frac{\partial I_{ds}}{\partial V_G}\right|_{V_D} , \quad \text{whence } g_m r_{ds} = \mu_A$$

and one finds that*

$$g_m/g_{ds} = \mu_A = C_{gc}/C_{dc} , \tag{17}$$

where $$C_{dc}/C_{gc}{}^{*} = \frac{\text{effective drain-to-channel coupling capacitance}}{\text{gate to channel capacitance}} \cong \frac{T_{ox}\epsilon_{si}}{L\epsilon_{ox}} \tag{18}$$

and ϵ_{si} = dielectric constant of the intrinsic silicon substrate.

Therefore, the transistor amplification factor μ_A is dependent solely on the geometry. This result is particularly interesting since it is exactly analogous to that which one obtains for the vacuum tube triode;[9]

$$\mu \equiv \left.\frac{\partial V_P}{\partial V_G}\right|_{I_P} = \frac{C_{pk}}{C_{kg}} . \tag{19}$$

This again illustrates the strikingly close parallel between this type of transistor and the vacuum tube.

Since* $$g_m = G_o \frac{V_G - V_P}{V_P}$$

(G_o = unmodulated channel conductance) (20)

for the thick oxide case, we may put Eq. (17) in the following form:

$$g_{ds}/G_o = \left(\frac{C_{dc}}{C_{gc}}\right)\left(\frac{V_g - V_P}{V_P}\right) \tag{21}$$

*See Appendix I.

Hence, this theory predicts that for a fixed geometry and for substrate dopings meeting the stated assumptions, the normalized saturation drain conductance is a linear function of the normalized gate voltage, with a slope equal to an effective capacitance ratio dependent *only* on the geometry.

2. High-Field Carrier Multiplication

It can be seen that for sufficiently small dimensions, internal fields may become large enough to cause avalanche breakdown within the channel or even directly from the drain to the substrate. Breakdown in the channel will usually commence in the region of maximum field (i.e., the space-charge limited region near the drain). However, the drain voltage required for breakdown is not a monotonic function of the gate voltage. As the gate is made more positive to enhance the conductivity of the channel, the length of regions 2 and 3 decreases. This yields an increased electric field in these regions resulting in a diminishing of the excess drain voltage required for channel breakdown. Conversely, for a negative gate voltage, Region 1 vanishes and Regions 2 and 3 are enlarged yielding an increase in the drain voltage required for breakdown. Now, however, as the gate voltage continues to go negative significantly beyond cutoff, the gate field begins to add directly to and enhance the effect of the drain field, thereby again reducing the drain voltage required for channel breakdown.

It is to be expected that the shape of the channel breakdown curves will differ from those of the conventional avalanche breakdown diode. As carrier multiplication commences, a plasma is generated which tends to shield the effect of the gate electrode. (This is exactly analogous to the sheath formed around probes in a conventional gas plasma.) In the depletion mode this is a negative feedback effect, allowing the drain current to rise and hence, acting to "soften" the knee of the breakdown curve. Hence, one would expect the breakdown in the channel to be easily differentiated from the direct drain to substrate breakdown, since the latter is a conventional reverse-biased p-n junction diode. This "soft" channel avalanche yields a poor drain saturation resistance for units whose dimensions are small enough to exhibit this effect.

B. THERMAL EQUILIBRIUM IN TWO-TERMINAL MOS DEVICES

Analyses have been made[5,6] of MOS diodes and MOS triodes assuming the ideal case of a perfectly insulating oxide (although including the effect of surface states), and applying Boltzmann statistics to determine the characteristics of the space charge region at the oxide-semiconductor interface. The application of Boltzmann statistics, or any equilibrium statistics for that matter, rests upon the assumption that the system is in a state of thermal equilibrium, or at least so close to it as to be treated as such. The assumption of thermal equilibrium must be reconsidered carefully if one generalizes the model to include the effect of power input to the system. In the case of a two-terminal device such as the MOS diode, this can occur for A.C. operation or through a leaky oxide. However, in a three-terminal device such as the field-effect transistor, this power input is inherent in its basic mode of operation.

From the ideal theory, it is usually approximated that the surface potential, V_i, at the oxide-semiconductor interface of a MOS diode cannot exceed 0.75 volt[5,6] (with the body of the semiconductor taken as reference). At this point, the Boltzmann factor exp (V_i/kT) is sufficiently large so that any further increase in gate voltage is met with an increasing accumulation layer at the interface. For example, for an n-type semiconductor, the depletion layer ceases to grow, and further control voltage increases are taken up by an accumulation layer of holes. However, if means are now available to remove these holes (assuming an n-type body for convenience), such as transverse sweep fields or the aforementioned leaky oxide, then careful consideration must be given to the source of supply of the holes, and its rate limitations, if any. If the source can supply holes at a rate significantly in excess of their removal, then the use of Boltzmann statistics as an approximation is usually valid.

If not, then these statistics do not provide an accurate picture, and the steady-state configuration must be determined as a supply-flow limited process.

As an illustration of the above discussion, consider the "induced-channel" and depletion" type transistors, respectively. In the former type unit, with n-type source and drain contacts and a p-type body, the gate is made positive so as to deplete first the holes and finally cause an accumulation (inversion) layer of electrons. This accumulation layer inverts the surface, bridges the n-type source-drain contacts, and allows ohmic current flow. In this case, the electrons are supplied by the source contact, which may be considered an infinite supply. Hence, the use of Boltzmann statistics to predict the behavior of this unit is justified.[6]

In the depletion type unit, the gate is made negative so that electrons are depleted from the n-type channel. Assume for the moment that the surface potential at the silicon-oxide interface, V_i, reaches the value required (by Boltzmann statistics) to cause inversion *before* the channel has been completely depleted. The statistical theory then says that all further control voltages will be taken up by the accumulation layer of holes. If the field pattern of a long, shallow channel is considered, it can be shown* that the transverse field at the interface is approximately the same as that in the ohmic region of the channel for $V_D \gg V_i$. Thus, the hole accumulation layer at the surface will, in general, be subjected to a field which will sweep the holes to the source contact.

Whether or not the accumulation layer is destroyed depends solely on whether or not the source of the holes can supply them as quickly as the sweep field removes them. In the case of the typical electron conduction type unit, the holes are supplied by thermal generation and diffusion from the n-type channel and the depletion region underlying the oxide. (It will be assumed that the p-type substrate is reverse-biased with respect to the channel.) A typical value of this

*From the requirement that $\nabla \times E = -\partial B/\partial t = 0$.

thermal "back current" for a diode (considering junction generation-recombination as the dominant mechanism) is of the order of 10^{-5} to 10^{-7} amp/cm^2.[10] For units with channel lengths of the order of a few mils or less, a further calculation shows that a sweep field in the range of mv/cm is sufficient to remove these holes quickly enough to reduce the steady-state inversion layer significantly below the thermal equilibrium value.*

Since transverse fields in these transistors are typically of the order of $10^3 - 10^4$ volts/cm, it may theoretically be predicted that saturation and pinch-off (both of which would be prevented by an accumulation layer) will, in fact, be observed for all such transistors. It should be added that an oxide that is sufficiently leaky may also be considered (in terms of accumulation layer removal) to act more as a reverse-biased p-n junction than as an ideal MOS structure.

Experimental evaluation of the previous discussions appears in the following section.

*See Appendix II.

IV. EXPERIMENTAL RESULTS

A. SATURATION RESISTANCE

Several units were fabricated using 1000 ohm-cm p-type material as a substrate. A circular geometry was employed as shown in Fig. 11. The physical dimensions of the units were identical:

Channel length = 5 mils
Channel width = 50 mils
Oxide thickness = 2000 Å

The varying parameters are the unmodulated channel resistance, $1/G_o$, and the pinch-off voltage, V_P.

	Unit 1	Unit 2	Unit 3
Pinch-off Voltage, V_P	-2.5v	-4.0v	-20.0v
Channel Resistance, $1/G_o$	10.8K	18.5K	1.4K

A plot of the normalized saturation drain conductance, g_{ds}, versus gate voltage appears in Fig. 12. For units 1 and 2, there is a very good degree of linearity and the two curves agree within a few per cent. For unit 3, there is still "factor of two" agreement, which is quite good considering the "order of magnitude" variation in pinch-off voltage and channel resistance and the relative

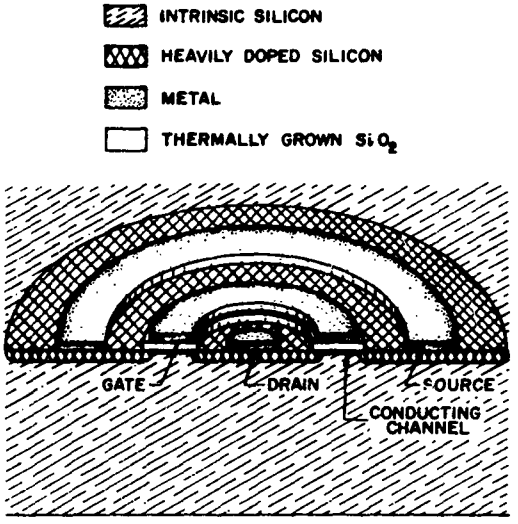


Fig. 11. Circular geometry for MOS field-effect transistors.

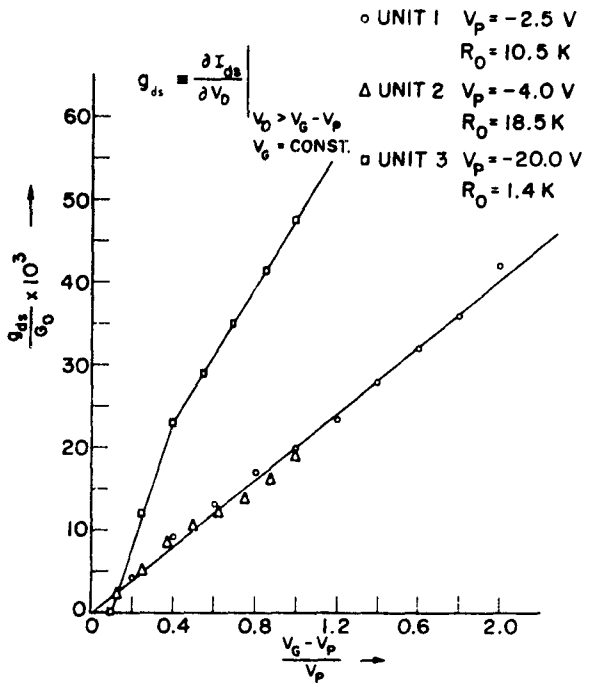


Fig. 12. Plot of g_{ds}/G_o vs $\frac{V_G - V_P}{V_P}$

simplicity of the assumed model. Calculating the slope ($C_{dc}/C_{gc} = 1/\mu_A$ from Eq. (18)) yields 0.5×10^{-2}. The observed values from Fig. 12 vary between 2.0×10^{-2} to 4.5×10^{-2}. Hence, the observed μ_A of the device was somewhat lower than the predicted value.

From the preceding results, it may be concluded that the proposed model does, in fact, provide a good representation for the behavior of these units in the saturation mode.

B. AVALANCHE BREAKDOWN

The breakdown characteristics predicted by the preceding theory were evidenced by all depletion type units tested (including those of noncircular geometry). A typical set of curves appears in Figs. 13 and 14. The expected "soft" breakdown at zero bias is quite apparent, with a "hardening" of the knee as gate voltage approaches pinch-off; the asymptote is the voltage for direct breakdown from drain to substrate. As the gate voltage is advanced past the pinch-off value, the drain voltage required for breakdown is now observed to decrease. It is interesting to note in Fig. 14 that the difference between the drain voltage and the gate voltage at the commencement of avalanche is fairly constant, tending to indicate that the "breakdown" occurs directly between the drain and gate with the substrate acting as a collector of the impact-ionized carriers.

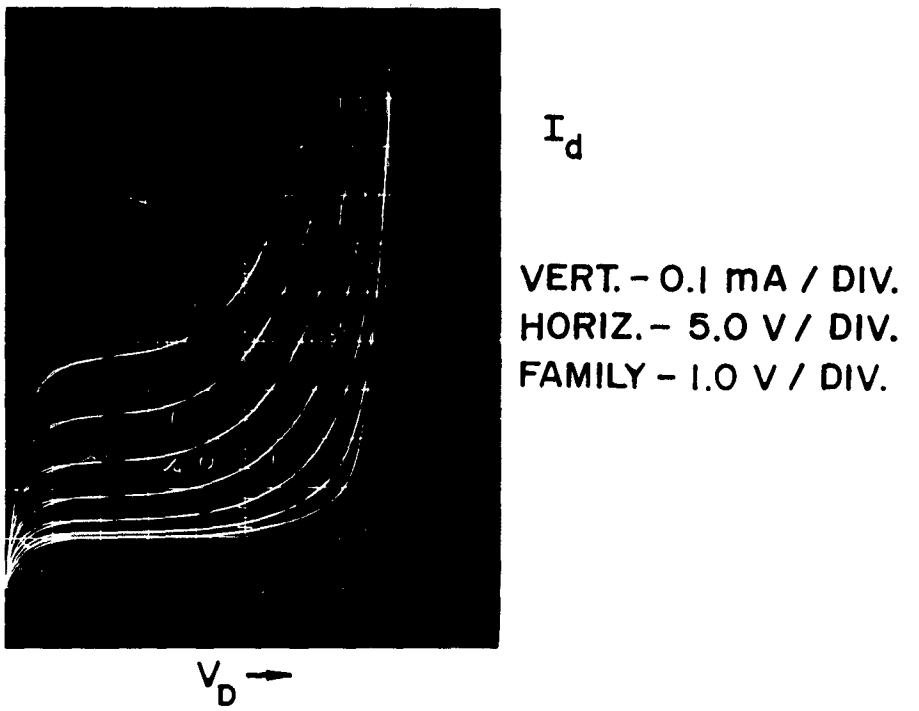


Fig. 13. Characteristic curves illustrating breakdown – n-type channel; $V_{gate} > V_{pinch\text{-}off}$.

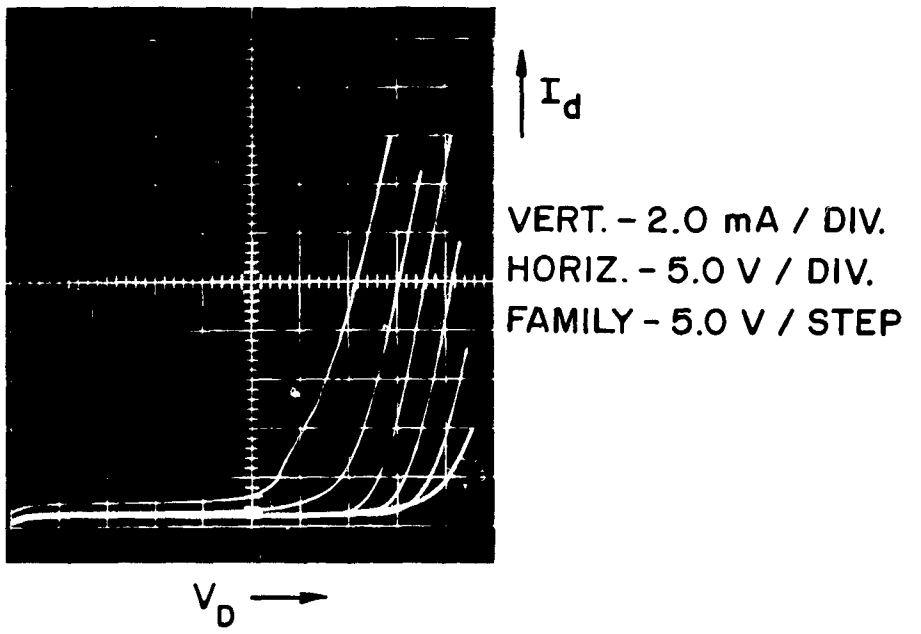


Fig. 14. Characteristic curves illustrating breakdown – n-type channel; $V_{gate} < V_{pinch\text{-}off}$.

C. INCOMPLETE PINCH-OFF

All units tested possessed saturation characteristics; non-zero currents at pinch-off were traced to reverse leakage between the drain and substrate due to poor diode characteristics. Hence, no evidence to support the presence of an inversion layer in depletion type units was found, even for units with 20 to 30 volt pinch-off voltages.

V. INTEGRATED CIRCUITS

Before choosing a logic circuit suitable for integration with induced channel field-effect transistors, the problems of yield and tolerance must be solved. It has been experimentally observed that the prime failure mechanism in these devices is a faulty gate insulator, probably caused by pinholes in the thermally grown oxide layer, yielding a short circuit between gate and channel. Direct measurements indicate that the oxide defects are randomly distributed on the surface of the silicon crystal and are most probably caused by irregular oxide growth in the neighborhood of surface imperfections. Assuming that the area of the pinhole is much smaller than the area of the gate electrode, one may postulate a model of randomly distributed points. Then, the probability of covering k pinholes with an electrode of area A is given by the Poisson distribution,

$$P(k; A) = \frac{(nA)^k}{k!} e^{-nA} , \tag{22}$$

where n is the average surface density of pinholes. The probability of success, P_s, is just the probability of covering zero pinholes and is given by,

$$P_s = P(0, A) = e^{-nA} . \tag{23}$$

Since the pinholes were assumed to be randomly distributed, the success of one unit is independent of the success of an adjacent unit (uncorrelated events), and the probability of obtaining N good transistors in an array is simply,

$$P_c \equiv (P_s)^N = (e^{-nA})^N = e^{-n(AN)}, \tag{24}$$

and depends only on the total active area of the circuit, NA. The pinhole density, n, has been determined experimentally by verification of Eq. (23); its value is approximately 1000 cm^{-2} for the crystals that were investigated. Thus, a transistor having a gate electrode area of $6 \times 1 \times 10^{-6}$ $in.^2 = 3.87 \times 10^{-5}$ cm^2, has a probability of success equal to $P_s = 0.968$, and an array of 16 such transistors has a probability of success equal to $P_c = 0.538$; these figures have been confirmed on a series of recently fabricated wafers. When room for wiring and dicing is taken into consideration, a normal 1-in. diameter silicon wafer will easily accommodate 20 circuits containing a 4×4 array of 16 transistors per circuit. The probability of obtaining exactly m operative circuits out of 20, with a probability of success for each circuit given by P_c is simply

$$P_m = \frac{(20)!}{m!(20-m)!} P_c^m (1 - P_c)^{20-m} \tag{25}$$

The probability of getting at least r operative circuits is the sum of probabilities, P_m, for $r \leq m \leq 20$;

$$P_{m\geq r} = \sum_{m=r}^{20} \frac{(20)!}{m!(20-m)!} P_c^m (1-P_c)^{20-m} . \qquad (26)$$

The plot of $P_{m\geq r}$ as a function of r, (with $P_c = 0.5$ for convenience) is shown in Fig. 15.

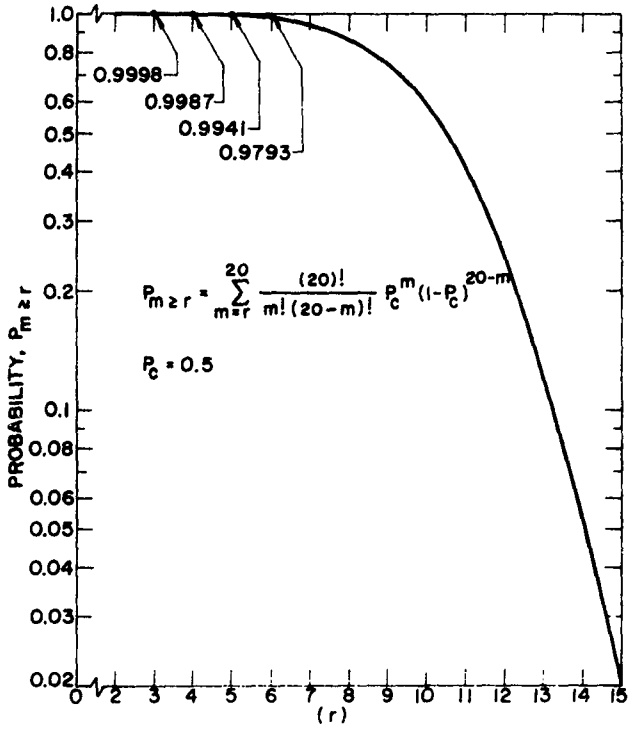


Fig. 15. Probability of obtaining at least r operative circuits out of 20.

Although the overall yield of operative circuits is 50 per cent, there may be occasions when a certain minimum number of operative circuits per wafer becomes the governing parameter. As an example, consider the special case where it is desired to integrate r circuits from the 20 on each wafer, and furthermore, assume that any pattern of the r circuits may be integrated just as easily. The total yield of useful wafers is then given by the factor $P_{m\geq r}$. If, for instance, it is desired to integrate five circuits out of the 20, then Fig. 15 gives a total yield of successfully integrated wafers in excess of 99.5 per cent.

In the light of the above results, a logic block containing 16 transistors was fabricated; the schematic diagram is shown in Fig. 16. The necessary load resistors will be fabricated as a part of the package, and will not be integrated on the silicon wafer. Figure 17 is a photograph of the wiring layout on the surface of the silicon wafer. All wiring is insulated from the silicon by evaporated silicon monoxide and room is left between circuits for dicing. A change in circuit design only necessitates a new wiring pattern; the transistor layout shown in Fig. 9 is fabricated before wiring.

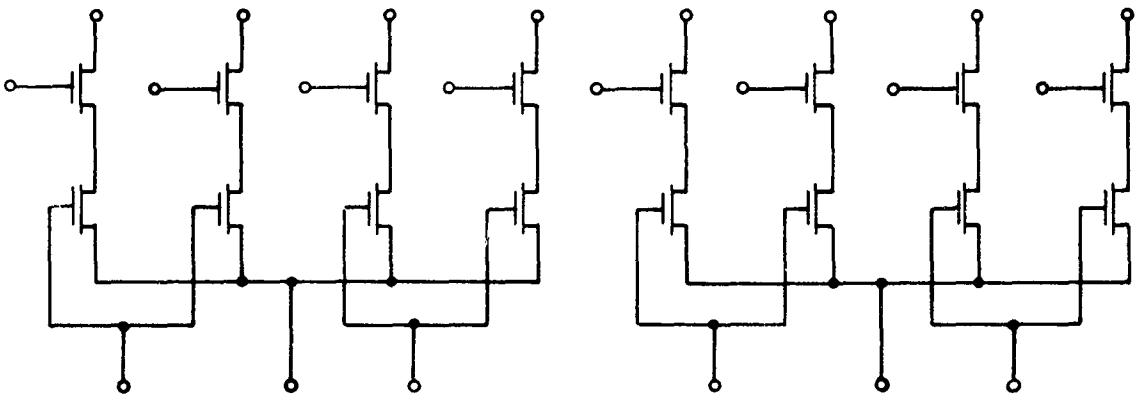


Fig. 16. Schematic diagram of 16-transistor logic block.

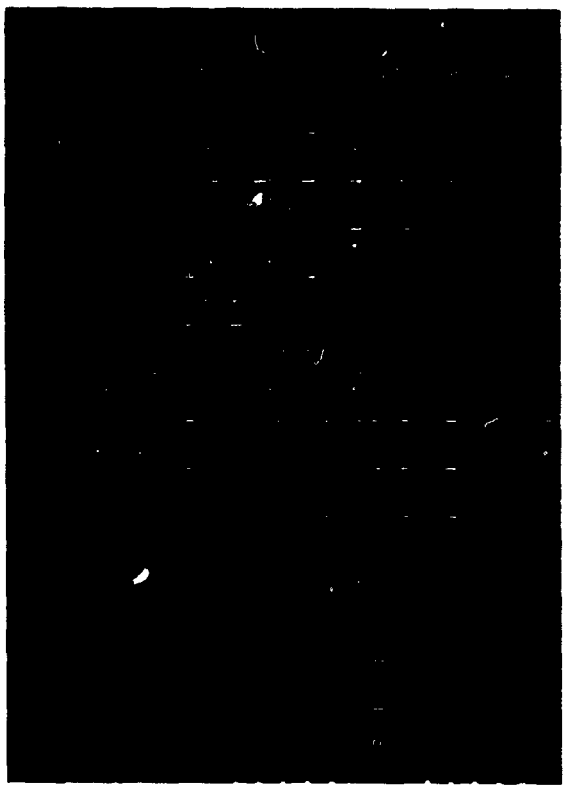


Fig. 17. Wiring layout of integrated logic blocks.

The tolerance problem, when applied to logical circuits, reduces to a problem of worst case design; i.e., all transistors having at least a minimum performance parameter are acceptable. The circuit analyzed is the simple logical inverter shown in Fig. 18, where the parasitic resistances, R_S and R_D, of the induced channel field-effect transistor have been assumed zero. (This assumption is quite good in the recently fabricated devices.) The condition for stable inversion is that the output voltage of the inverter, V_D, be less than V_P (a positive quantity for the n-p-n induced channel unit), when the input voltage is V_B, the supply voltage. This means that V_D applied to a successive stage will cut off this transistor and produce an output of V_B. Equating V_D to V_P gives the equations for marginal stability.

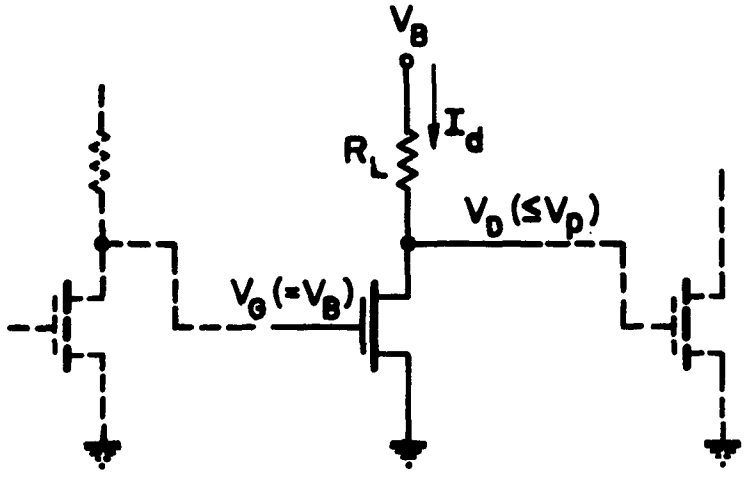


Fig. 18. Circuit of simple logical inverter.

With $R_S = R_D = 0$, Eq. (8) becomes

$$I_d = \beta(V_G - V_P)V_D\left[1 - \frac{1}{2}\,\frac{V_D}{V_G - V_P}\right], \tag{27}$$

and expresses the drain characteristics of the transistor before saturation of drain current. If the load resistance, R_L, is chosen to operate in this region, we have, setting $V_D = V_P$ and $V_G = V_B$,

$$I_d = \beta(V_B - V_P)V_P\left[1 - \frac{1}{2}\,\frac{V_P}{V_G - V_P}\right] \tag{28}$$

and

$$V_B - V_P = I_d R_L \ . \tag{29}$$

Eliminating I_d from the above equations, yields the following relation between R_L, V_B, β, and V_P:

$$R_L = \frac{1}{(\beta V_P)} \frac{1}{\left[1 - \frac{1}{2(V_B/V_P - 1)}\right]} \tag{30}$$

From physical reasoning, it is seen that the load resistor given above is the minimum value that can be tolerated to yield logical stability. The function multiplying $1/\beta V_P$ is monotonically decreasing for $V_B/V_P > 1$, and lies between 2 and 1 for $V_B/V_P \geq 2$. Therefore, for logical stability, with a reasonable choice of supply voltage, V_B,

$$R_L \geq \frac{2}{\beta V_P} \quad . \tag{31}$$

The worst case design dictates that we choose a load resistor compatible with a transistor having the smallest β and V_P. If we assume that the output of each transistor is connected to the gate of f others (f is the fan-out), then $\tau = fR_L C_{in}$ is the time constant of each stage. Using the equality in Eq. (31) yields

$$\tau = fR_L C_{in} = f \frac{2L^2}{\mu V_P} \quad , \tag{32}$$

where τ is roughly the rise time, or pair delay of the inverter. This analysis was carried out for the n-p-n induced channel field-effect transistor, where V_P is a positive number.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the able assistance of G. A. Brown and R. R. Vannozzi in fabricating these devices and alleviating many of the practical difficulties encountered. The information obtained from J. Olmstead and J. Scott of RCA, Somerville, on the diffusion and cleaning procedures was essential to the realization of this transistor. Many important points were brought to light in the stimulating discussions with T. O. Stanley and W. M. Webster. We would also like to thank J. McCusker, L. Pensak, A. Rose, M. E. Sekely, R. B. Schilling, H. S. Sommers, Jr., J. T. Wallmark, and K. H. Zaininger for their helpful comments. J. Briggs is responsible for the delicate job of dicing the wafers and G. Lang did the chemical polishing.

APPENDIX I

For the following it will be assumed that:

a) The oxide thickness > 3 × channel depth.
b) The silicon substrate is intrinsic and may be treated as an insulator.
c) The transistor is operating in the saturation region.
d) All quantities are in terms of per unit width of the channel unless otherwise specified.

Condition a) assures that most of the gate-channel voltage drop appears across the oxide. This yields a square-law characteristic – i.e.,

$$g_m \equiv \left.\frac{\partial I_{ds}}{\partial V_D}\right|_{V_G} = G_0 \frac{V_G - V_P}{V_P} \tag{I-1}$$

where $G_0 \equiv \left.\frac{\partial I_d}{\partial V_D}\right|_{\substack{V_G = 0 \\ V_D = 0}}$ = unmodulated channel conductance.

A change in drain voltage ΔV_D will therefore cause a change in the average channel sheet charge of $\Delta Q = C_{dc}\ \Delta V/L$ per unit area:

where C_{dc} = effective drain-to-channel coupling capacitance, per unit width

L = channel length.

This change in charge sees an average electric field from source to drain of $\frac{V_G - V_P}{L}$ with an effective channel mobility of μ_{eff}. The change in the drain current may therefore be written as:

$$\Delta I_{ds} = \left(\frac{C_{dc}\Delta V_D}{L}\right)\left(\frac{V_G - V_P}{L}\right)(\mu_{eff}) . \tag{I-2}$$

or

$$g_{ds} = \left.\frac{\Delta I_{ds}}{\Delta V_D}\right|_{V_G} = \frac{C_{dc}(V_G - V_P)\mu_{eff}}{L^2} \tag{I-3}$$

The above relation may be further simplified by use of the following two subrelations:

$$G_0 = Q\mu_{eff}\ L\ , \tag{I-4}$$

$$V_P = \left(\frac{L}{C_{gc}}\right)Q\ , \tag{I-5}$$

where Q = unmodulated channel sheet charge/unit area

and C_{gc} = gate to channel capacitance.

Using these relations in Eq. (I-3) yields:

$$g_{ds} = g_m (C_{dc}/C_{gc}) . \tag{I-6}$$

Since $C_{cg} = (L/T_{ox})\ (\epsilon_{ox})$, Eq. (I-6) may be rewritten as

$$g_m/g_{ds} = \mu_{in} = C_{gc}/C_{dc} = \frac{L}{T_{ox}} \frac{\epsilon_{ox}}{\epsilon_{si}} . \tag{I-7}$$

It should be mentioned at this point, that C_{dc} is an equivalent "lumped" feedback capacitance; actually, the drain-to-channel feedback capacitance is distributed along the channel length, and its effect on the channel charge and drain current must be computed using an incremental technique similar to the one used to calculate the basic device characteristic equations. However, as a first approximation, the model of Fig. 10 may be used with the assumption that the field lines between the drain and the underside of the channel are semicircles. Then the incremental capacitance from the drain to a section of channel dx in length is simply

$$d(C_{dc}) = \frac{\epsilon_{si}}{\pi x} dx \tag{I-8}$$

where x = the distance from the edge of the drain contact to dx. To obtain the total capacitance the above expression must be integrated over the ohmic regions of the channel (Regions 1 and 2).

$$C_{dc} = \int_{X_{min}}^{L} \frac{\epsilon_{si}}{\pi x} dx = \frac{\epsilon_{si}}{\pi} \ln (L/X_{min}) . \tag{I-9}$$

X_{min} corresponds to the beginning of Region 3 (the space-charge region) and hence will be a function of V_d. However, C_{dc} varies logarithmically with X_{min} and hence will be relatively invariant. Setting L/X_{min} typically at $10^{+1} - 10^{+2}$ yields

$$C_{dc} \approx \epsilon_{si} ; \text{ unit width of channel.} \tag{I-10}$$

APPENDIX II

Assume an n-type channel and hence an inversion layer of holes.

Let:

I_r = reverse saturation hole current to gate (assuming for the moment that the gate is a hole sink; e.g., a p-n junction)

E_s = transverse sweep field,

L_c = channel length,

μ_e = effective mobility of the holes in the inversion layer,

T_{ox} = oxide thickness,

Qe_{TH} = inversion layer density (sheet charge/unit area) from the thermal equilibrium equation,

Qe_{ss} = the actual steady-state inversion layer density,

ϵ_{ox} = dielectric constant of the SiO_2,

V_{Gi} = gate potential for which the surface potential V_i (at the oxide-silicon interface) is just sufficient to cause inversion.

From the requirement of conservation of charge, one may then write to a good approximation

$$Qe_{ss} E_s \mu_e \approx I_r L_c \ ,$$

whence

$$Qe_{ss} = \frac{I_r L_c}{E_s \mu_e} \ .$$

Now

$$Qe_{TH} = \left(\frac{V_G - V_{Gi}}{T_{ox}} \right) (\epsilon_{ox}) \quad \text{(for a thick oxide)} \quad V_G > V_{Gi}$$

and hence

$$Qe_{ss} / Qe_{TH} = \frac{(I_r L_c)(T_{ox})}{(E_s \mu_e)(V_G - V_{Gi})\epsilon_{ox}} \ .$$

For a typical unit

$L_c \approx 10^{-3}$ cm (worst case approximation)

$\mu_e \approx 300$ cm²/v- sec

$T_{ox} \approx 1000$ Å $= 1 \times 10^{-5}$ cm,

$\epsilon_{ox} \approx \frac{1}{3} \times 10^{-12}$ f/cm,

and letting, typically, $V_G - V_{Gi} = 10$ v, one has

$$Qe_{ss} / Qe_{TH} = \left(\frac{10^{-5} \times 10^{-3}}{300 E_s} \right) \left(\frac{1}{(1/3) \times 10^{-12}} \right) \left(\frac{10^{-5}}{10} \right) \approx \frac{10^{-4}}{E_s} \ .$$

Hence, $E_s \geq$ mv/cm is adequate to reduce the inversion layer far below its thermal equilibrium value.

Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, Massachusetts, AFCRL-63-331, INTEGRATED LOGIC NETS, Scientific Report No. 2, 34 p. incl. illus., and 10 refs., July 1963.

Unclassified Report

This report describes a device possessing a significant departure from the conventional unipolar transistor in that the reverse-biased p-n junction has been superseded by a metal-oxide-semiconductor (MOS) control structure. A simple model is proposed and the basic transistor current-voltage relationships are derived for a thick oxide and shallow conducting channel. A more detailed model is then proposed to explain some observed anomalies. The usual approximation of constant current in the saturation region is abandoned and the behavior of the drain resistance is considered. Relations predicting and describing this behavior are closely analogous to similar relations for vacuum tubes. Experimental data from units to which this model may be applied have shown close agreement with the theoretical predictions.

The validity of using Boltzmann or Fermi-Dirac statistics in devices not necessarily in thermal equilibrium is discussed and found to be unjustified in some cases.

The yield of units fabricated to date has averaged over 95% on recently fabricated wafers, indicating great promise for integrated electronics applications.

1. Integrated circuits
2. Field-effect transistors
3. Semiconductor device
4. Computer logic

I. AFSC Project 4641, Task 464104
II. Contract No. AF19(604)-8836
III. RCA Laboratories, Princeton, New Jersey
IV. S.R.Hofstein and F.P. Heiman
V. In DDC collection

Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, Massachusetts, AFCRL-63-331, INTEGRATED LOGIC NETS, Scientific Report No. 2, 34 p. incl. illus., and 10 refs., July 1963.

Unclassified Report

This report describes a device possessing a significant departure from the conventional unipolar transistor in that the reverse-biased p-n junction has been superseded by a metal-oxide-semiconductor (MOS) control structure. A simple model is proposed and the basic transistor current-voltage relationships are derived for a thick oxide and shallow conducting channel. A more detailed model is then proposed to explain some observed anomalies. The usual approximation of constant current in the saturation region is abandoned and the behavior of the drain resistance is considered. Relations predicting and describing this behavior are closely analogous to similar relations for vacuum tubes. Experimental data from units to which this model may be applied have shown close agreement with the theoretical predictions.

The validity of using Boltzmann or Fermi-Dirac statistics in devices not necessarily in thermal equilibrium is discussed and found to be unjustified in some cases.

The yield of units fabricated to date has averaged over 95% on recently fabricated wafers, indicating great promise for integrated electronics applications.

1. Integrated circuits
2. Field-effect transistors
3. Semiconductor device
4. Computer logic

I. AFSC Project 4641, Task 464104
II. Contract No. AF19(604)-8836
III. RCA Laboratories, Princeton, New Jersey
IV. S.R.Hofstein and F.P. Heiman
V. In DDC collection